DICTIONARY LEARNING FOR POISSON COMPRESSED SENSING

Sukanya Patil, Rajbabu Velmurugan

Dept. of Electrical Engineering, IIT Bombay
{sukanya_patil,rajbabu}@ee.iitb.ac.in

Ajit Rajwade

Dept. of CSE, IIT Bombay
ajitvr@cse.iitb.ac.in

ABSTRACT

Imaging techniques involve counting of photons striking a detector. Due to fluctuations in the counting process, the measured photon counts are known to be corrupted by Poisson noise. In this paper, we propose a blind dictionary learning framework for the reconstruction of photographic image data from Poisson corrupted measurements acquired by a *compressive* camera. We exploit the inherent non-negativity of the data by modeling the dictionary as well as the sparse dictionary coefficients as non-negative entities, and infer these directly from the compressed measurements in a Poisson maximum likelihood framework. We experimentally demonstrate the advantage of this *in situ* dictionary learning over commonly used sparsifying bases such as DCT or wavelets, especially on color images.

***Index Terms*—** Poisson compressed sensing, non-negative sparse coding, dictionary learning

1. INTRODUCTION AND RELATION TO PRIOR WORK

Compressed sensing (CS) is a widely used paradigm to acquire sparse signals using a small number of incoherent observations of the signal. There exist several algorithms for reconstruction from compressed measurements, a majority of which operate in the noiseless regime or assume Gaussian or bounded uniform noise, *e.g.*, [1, 2, 3], etc. However, imaging techniques involve counting of photons striking a detector array and the random fluctuations in photon counts are known to follow a Poisson distribution. These fluctuations cannot be modeled using signal independent constant variance models such as Gaussian noise, especially in the low light or low photon count regime, which is common in fluorescence microscopy, low dose tomography and night photography. Due to the signal dependent and non-additive nature of Poisson noise, compressed sensing from Poisson-corrupted measurements poses significant challenges, which have been carefully studied in [4] along with the derivation of bounds on reconstruction errors. Assuming fixed bases such as wavelets to provide a sparsifying transform, a reconstruction algorithm from Poisson-corrupted measurements named SPIRAL-TAP has been proposed in [5]. Non-monotonic maximum likelihood algorithms in conjunction with maximum entropy priors on the image intensity (instead of sparsity in a basis) have been proposed in [6] for applications such as image deblurring, transmission tomography and PET reconstruction.

In compressed sensing, the sparsity of a signal of interest in a certain pre-specified sparsifying basis, has great importance. In particular, the performance of Poisson CS reconstruction is shown to be highly dependent upon the chosen sparsifying basis [4, 7]. This provides us the motivation to develop an algorithm to infer a sparsifying basis *in situ* from the compressed measurements directly. This is called as blind compressed sensing, and has been explored in the Gaussian noise case in works such as [8, 9, 10]. There also exists a substantial body of literature on inference of data-adaptive dictionaries from Poisson-corrupted images, *e.g.*, [11, 12, 13, 14, 15]. Besides incorporating a criterion to encourage signal sparsity in the inferred dictionary, these techniques are based on a Poisson maximum likelihood framework. They pose a viable alternative to techniques such as [16] that (a) seek to convert Poisson-corrupted image data to Gaussian-corrupted data using variance-stabilizing transforms (VSTs) such as the Anscombe transform, (b) perform inference of dictionaries and coefficients using least-squares techniques that are suited to Gaussian noise, and (c) then convert the processed data back to the original domain using the inverse VSTs. The disadvantage of the latter group of techniques is that the Gaussian approximation of the noise following the VST is inaccurate at low photon counts requiring more careful implementation of the inverse VST [17].

In this paper, we propose a simple and efficient algorithm to infer a sparsifying dictionary and the dictionary coefficients *in situ* from Poisson-corrupted *compressive measurements* in imaging (as opposed to Poisson-corrupted images), instead of using standard bases such as wavelets or DCT. To the best of our knowledge, there exists no prior literature on this so far. We experimentally demonstrate that learning a dictionary particular to color images allows for better reconstruction than using standard sparsifying bases.

2. PROBLEM FORMULATION

Consider an underlying noise-free image $\mathbf{X} \in \mathbb{Z}^{N_1 \times N_2}$. Consider a division of the image into patches of size $n_1 \times n_2$, with the i^{th} patch represented as a vector $\mathbf{x}_i \in \mathbb{Z}^{n\times 1}$ where $n \triangleq n_1 n_2$. Consider that a compressive sensor acquires $m \ll n$ measurements of each patch $\mathbf{x}_i$ to produce a measurement vector $\mathbf{y}_i \in \mathbb{Z}^m$. We assume that each such measurement vector is corrupted by Poisson noise, giving us the following equation:

$$\forall i, 1 \leq i \leq N_p, \mathbf{y}_i \sim \text{Poisson}(\mathbf{\Phi}_i \mathbf{x}_i) \tag{1}$$

where $\mathbf{\Phi}_i \in \mathbb{Z}^{m\times n}$ is the forward model matrix of the measuring device for the i^{th} patch, and N_p is the total number of patches. The original vectors $\mathbf{x} \triangleq \{\mathbf{x}_i\}_{i=1}^{N_p}$ are assumed to be sparse in some dictionary $\mathbf{D}$ with some K columns each of unit norm, i.e. we have $\mathbf{x}_i = \mathbf{D}\mathbf{s}_i$ where $\|\mathbf{s}_i\|_0 \ll n$. The aim here is to infer both $\mathbf{D}$ and $\mathbf{s} \triangleq \{\mathbf{s}_i\}_{i=1}^{N_p}$ from $\mathbf{y} \triangleq \{\mathbf{y}_i\}_{i=1}^{N_p}$. During this inference, we impose non-negativity constraints on the elements of the dictionary $\mathbf{D}$ (in addition to unit norm constraints on its columns) as well as on the sparse coefficients $\mathbf{s}$. This is motivated by the fact that the underlying data are indeed non-negative. Taking this into account, we now have:

$$\mathbf{y}_i \sim \text{Poisson}(\mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i) \text{ s.t. } \mathbf{D} \succeq \mathbf{0}^{n\times K}; \forall i, \mathbf{s}_i \succeq \mathbf{0}^{K\times 1} \tag{2}$$

Assuming statistical independence between the different measurement vectors, we seek to maximize the following likelihood in order to infer $\mathbf{D}$ and $\mathbf{s}$:

$$\begin{aligned} \mathcal{L}(\mathbf{D}, \mathbf{s}) &= \prod_{i=1}^{N_p} p(\mathbf{y}_i | \mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i) \\ &= \prod_{i=1}^{N_p} \prod_{j=1}^{m} \frac{(\mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i)_j{}^{\mathbf{y}_{ij}}}{\mathbf{y}_{ij}!} e^{-(\mathbf{\Phi_i D s}_i)_j} \end{aligned} \tag{3}$$

where $\mathbf{y}_{ij}$ indicates j^{th} compressive measurement of i^{th} patch. Considering the negative log likelihood function along with a sparsity-promoting regularization term proportional to $\sum_i \|\mathbf{s}_i\|_q$ where $0 < q \leq 1$ (which is the negative logarithm of an appropriate Generalized Gaussian prior) and dropping constant terms, yields us the following objective function to be minimized:

$$\mathcal{J}(\mathbf{D}, \mathbf{s}) = \tag{4}$$

$$\min_{\mathbf{D},\mathbf{s}} \sum_{i,j} \bigg(-\mathbf{y}_{ij} \log((\mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i)_j) + (\mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i)_j \bigg) + \lambda \sum_{i=1}^{N_p} \|\mathbf{s}_i\|_q$$

subject to the constraints $\mathbf{D} \succeq \mathbf{0}^{n\times K}; \forall i, \mathbf{s}_i \succeq \mathbf{0}^{K\times 1}$ where λ is the regularization parameter. We set $q = 1$, though in principle any other q in the $(0, 1]$ range could have been used.

We use projected gradient descent with adaptive step size to minimize $\mathcal{J}(\mathbf{D}, \mathbf{s})$. The gradients of $\mathcal{J}(\mathbf{D}, \mathbf{s})$ with respect to $\mathbf{D}, \mathbf{s}$ are as shown below:

$$\begin{aligned} \forall i, 1 \leq i \leq N_p, \frac{\partial \mathcal{J}}{\partial \mathbf{s}_i} &= (\mathbf{\Phi}_i \mathbf{D})^T [\mathbf{1} - (\mathbf{y}_i./\mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i)] + \lambda \\ \frac{\partial \mathcal{J}}{\partial \mathbf{D}} &= \sum_{i=1}^{N_p} [\mathbf{\Phi_i}^T (\mathbf{1} - (\mathbf{y_i}./\mathbf{\Phi_i D s_i})] \mathbf{s_i}^T \end{aligned} \tag{5}$$

where $\mathbf{1}$ is a m-dimensional vector of ones and ./ denotes element-wise division. Instead of projected gradient descent which we have used in our algorithm, one can use modified multiplicative update to update every $\mathbf{s_i}$ [18]. Our algorithm is summarized here below.

Algorithm 1 Algorithm for reconstruction in Poisson CS

Input: Poisson corrupted data, $\mathbf{y}_i \sim \text{Poisson}(\mathbf{\Phi}_i \mathbf{D}\mathbf{s}_i), 1 \leq i \leq N_p$

Set $k = 0$. Randomly initialize $\mathbf{D}^{(0)}$ and $\mathbf{s}^{(0)}$ with appropriate sizes and non-negative entries. Set the ℓ_2 norm of all the columns of $\mathbf{D}$ to 1.
repeat
  while $\mathcal{J}(\mathbf{D}^{(k+1)}, \mathbf{s}^{(k)}) > \mathcal{J}(\mathbf{D}^{(k)}, \mathbf{s}^{(k)})$ **do**
    $\tilde{\boldsymbol{D}} = \mathbf{D}^{(k)} - \alpha_1 \frac{\partial \mathcal{J}}{\partial \mathbf{D}}\big|_{\mathbf{D}=\mathbf{D}^{(k)}}$
    Set all negative entries in $\tilde{\boldsymbol{D}}$ equal to zero
    Rescale each column of $\tilde{\boldsymbol{D}}$ to unit norm.
    Set $\mathbf{D}^{(k+1)} = \tilde{\boldsymbol{D}}$
    $\alpha_1 \leftarrow \eta \alpha_1$
  end while
  while $\mathcal{J}(\mathbf{D}^{(k+1)}, \mathbf{s}^{(k+1)}) > \mathcal{J}(\mathbf{D}^{(k+1)}, \mathbf{s}^{(k)})$ **do**
    $\tilde{\boldsymbol{s}} = \mathbf{s}^{(k)} - \alpha_2 \frac{\partial \mathcal{J}}{\partial \mathbf{s}}\big|_{\mathbf{s}=\mathbf{s}^{(k)}}$
    Set all negative entries in $\tilde{\boldsymbol{s}}$ equal to zero
    Set $\mathbf{s}^{(k+1)} = \tilde{\boldsymbol{s}}$
    $\alpha_2 \leftarrow \eta \alpha_2$
  end while
  $k = k + 1$
until stopping criteria is met

2.1. Modeling non-negativity:

We emphasize that one of the differences in Poisson CS and other CS paradigms is that signals considered in the former must inherently be non-negative as they emerge from a photon-counting process. Therefore, any estimate of the reconstructed signal needs to be non-negative. To impose the constraint $\mathbf{x}_i \succeq \mathbf{0}^{n\times 1}$, one may use either of the following three models: (i) $\mathbf{x}_i = \mathbf{D}\mathbf{s}_i$, s. t. $\mathbf{D} \succeq \mathbf{0}^{n\times K}, \mathbf{s}_i \succeq \mathbf{0}^{K\times 1}$ as used in this paper or in [12], (ii) $\mathbf{x}_i = \mathbf{D}\mathbf{s}_i$, s. t. $\mathbf{D}\mathbf{s}_i \succeq \mathbf{0}^{K\times 1}$ as used in [5], or (iii) $\mathbf{x}_i = \exp(\mathbf{D}\mathbf{s}_i)$ as used in [13, 11]. The advantage of model (i) over (iii) is that in the former case, we are representing the (raw intensity) patches from the image as non-negative linear combinations of the dictionary columns. Such a dictionary will always have a physical significance and represents basis patches like edges or patterns from the image. On the other hand, model (iii) would represent the element-wise logarithm of the patches as a non-negative linear combination, which is less intuitive. The advantage of model (i) over (ii) is simpler optimization, as in the former case, projection on the constraint sets simply involves clamping negative values from $\mathbf{D}$ or $\mathbf{s}_i$ to 0, whereas model (ii) does not afford that freedom.

Choice of $\mathbf{\Phi}$: The entries of $\mathbf{\Phi}$ must necessarily be non-negative since photon counts are being modeled. Moreover, under the model that *all* compressed measurements for a patch are acquired simultaneously in an optical system, it is known that the incident photon flux gets distributed amongst the different measurements. This is modeled by using non-negative $\mathbf{\Phi}$ matrices whose columns sum up to a value less than or equal to 1 [7]. The forward models chosen in this paper are matrices with entries sampled independently from $\text{Uniform}(0, 1)$ followed by a rescaling of the columns to sum up to some value (say) $v \leq 1$. For convenience of representation, these matrices can equivalently be considered without rescaling of the columns, but instead scaling down the peak value of the image by the appropriate factor to get the maximum column sum to be v.

3. EXPERIMENTS AND RESULTS

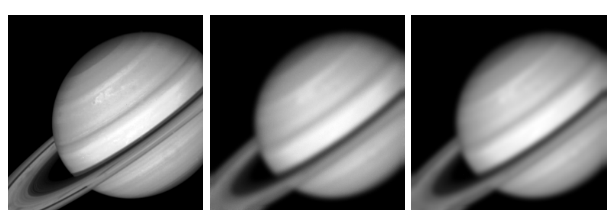
We present reconstruction results on gray-scale and color images under varying levels of compression and peak intensities of the image. The specific compressive forward model considered in this paper is based on computation of random dot products, which is similar to the functioning of the Rice Single Pixel Camera [19]. However, our measurements are computed on small patches and not on the entire image, which follows the camera architecture in [20]. The results are demonstrated on overlapping image patches, although our models are perfectly applicable for the case of non-overlapping patches, which is more in tune with the architecture of [20], but which would require additional preprocessing to overcome patch seam artifacts.

We performed reconstruction experiments on two different peak images values $\in \{3, 10\}$ to simulate low light scenarios (assuming the original image intensities lie in the [0,255] range) and three different compression levels given as $\frac{m}{n} \in \{0.2, 0.5, 0.8\}$. In all experiments, the number of dictionary columns K was set to 100, the patch size was set to 7×7 (vectorized to 49×1), and the value of λ was chosen from $\{0.1, 1, 5, 10, 20\}$ to be the one that produced the best result in terms of highest PSNR (peak signal to noise ratio). The iterative optimization was run for 100 iterations starting with random but feasible initial conditions. The proposed algorithm was compared to the well-known SPIRAL-TAP method [5] on two grayscale images - namely Saturn and House. A patch-based variant of SPIRAL-TAP was used for a fair comparison between these two algorithms. The sparsifying basis used for SPIRAL-TAP was the 2D-DCT for 7×7 patches

Image	Peak	Compression level ($\frac{m}{n}$)	Our Algorithm	SPIRAL-TAP
House	3	0.2	23.73	19.70
	3	0.5	24.07	20.91
	3	0.8	24.11	22.41
	7	0.2	23.73	19.70
	7	0.5	24.14	20.91
	7	0.8	24.15	22.42
	10	0.2	24.11	19.71
	10	0.5	24.12	20.91
	10	0.8	24.12	22.41
Saturn	3	0.2	29.48	22.41
	3	0.5	29.79	24.12
	3	0.8	29.82	26.54
	7	0.2	30.09	22.44
	7	0.5	30.19	24.10
	7	0.8	30.13	26.55
	10	0.2	30.25	22.43
	10	0.5	30.20	24.11
	10	0.8	30.25	26.55

Table 1. PSNR comparison for House and Saturn image under various compression levels and peak intensities

with a maximum of 100 iterations and λ chosen in the same manner as for our method. We also experimented with Haar wavelets on 8×8 patches and the results were not significantly different. We have displayed the PSNR values in Table 1 and the reconstructed images for visual comparison of the outputs for both these algorithms in Figures 1 and 2. While our method produced higher PSNR, the results are not significantly different in a visual sense. We also observed that reconstructed images from SPIRAL-TAP have a DC bias which results in the difference in PSNR. At much lower peaks, we have observed failure of our algorithm starting from a random initial condition due to lack of information (for dictionary and sparse code inference) available from compressed measurements in poor lighting. However using a pre-learned patch-based dictionary as initial condition may improve these results.



(a) Original image (b) Our algorithm, PSNR = 30.20 (c) SPIRAL-TAP, PSNR = 24.11

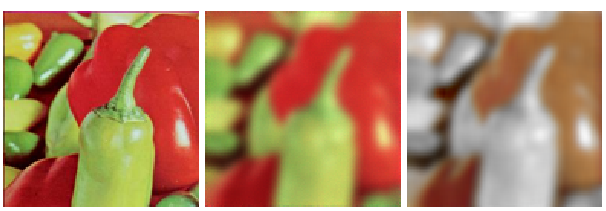
Fig. 1. Saturn image, Peak = 10 and Compression Ratio = 0.5

(a) Original image (b) Our algorithm, PSNR = 24.12 (c) SPIRAL-TAP, PSNR = 20.91

Fig. 2. House image, Peak = 10 and Compression Ratio = 0.5

3.1. Color Compressive Sensing:

We have generalized our algorithm for reconstruction of color images as well. Compressed measurements are computed independently across R, G and B channel with different forward model matrices, but the reconstruction is done jointly by learning a single dictionary over R, G, B channels using our algorithm. In this case, the image patches were of size $7 \times 7 \times 3$ (vectorized to 147×1) and $K = 150$ dictionary columns were learned. A similar experiment is performed with patch-based SPIRAL-TAP using 3D-DCT as a sparsifying basis. Figures 3, 4, 5 and 6 show reconstruction results for parts of three well-known color images using our method and SPIRAL-TAP. Here we observed that SPIRAL-TAP fails to reconstruct color patterns and produces color artifacts in the reconstructed images. This is because 3D-DCT is unable to compactly represent color image patches that contain pixels with significantly different R, G, B values. This motivates the idea of dictionary learning for multidimensional signals such as color or hyperspectral images, or videos, instead of relying on prior knowledge of a sparsifying basis.

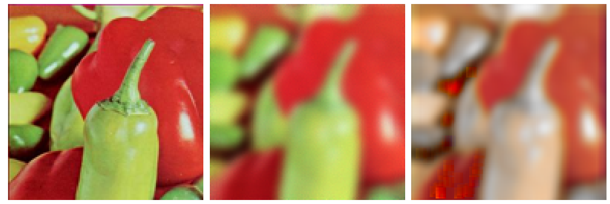


(a) Original image (b) Our algorithm, PSNR = 21.73 (c) SPIRAL-TAP, PSNR = 12.41

Fig. 3. Pepper image, Peak = 5 and Compression Ratio = 0.1

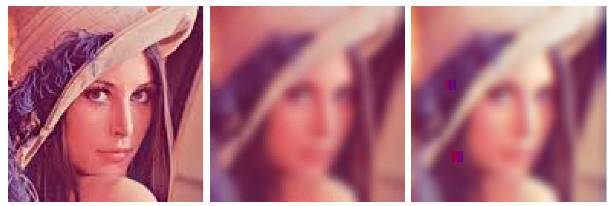
4. CONCLUSION AND FUTURE WORK

We have proposed a dictionary learning algorithm for reconstruction from Poisson-corrupted compressive measurements, for both gray-scale as well as color images. Our results illustrate the benefits of learning the dictionary *in situ* from the compressed measurements. While currently, our algorithm is able to recover global image structure and color patterns



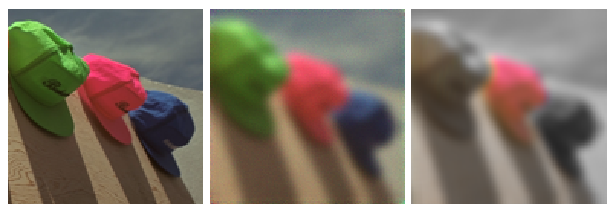
(a) Original image (b) Our algorithm, PSNR = 22.22 (c) SPIRAL-TAP, PSNR = 16.09

Fig. 4. Pepper image, Peak = 5 and Compression Ratio = 0.5



(a) Original image (b) Our algorithm, PSNR = 21.02 (c) SPIRAL-TAP, PSNR = 19.39

Fig. 5. Lena image, Peak = 5 and Compression Ratio = 0.5



(a) Original image (b) Our algorithm, PSNR = 25.63 (c) SPIRAL-TAP, PSNR = 20.72

Fig. 6. Cap image, Peak = 5 and Compression Ratio = 0.5

well, it tends to blur out finer textures. This effect has also been observed by us in denoising experiments on Poisson-corrupted images using state of the art algorithms such as non-local PCA [13] or non-negative sparse coding [12]. Furthermore, currently the choice of the λ parameter is unclear, and given the non-additive and signal-dependent nature of Poisson noise, cannot be tied to the noise variance in the same manner as Gaussian noise. In fact, for inappropriate values of λ, we have observed total failure of reconstruction using our dictionary learning method as well as SPIRAL-TAP. This issues requires further investigation. Exploring the application of this method for color image demosaicing, video and hyperspectral image reconstruction, and tomographic reconstruction are also interesting avenues for future research.

5. ACKNOWLEDGEMENT

Part of the work was supported by Bharti Centre for Communication in IIT Bombay.

6. REFERENCES

[1] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, “A method for largescale l1-regularized least squares,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[2] D. Needell and J. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, pp. 301–321, May 2009.

[3] T. Cai and W. Lie, “Orthogonal matching pursuit for sparse signal recovery with noise,” *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.

[4] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia, “Compressed sensing performance bounds under Poisson noise,” *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 3990–4002, Aug 2010.

[5] Z. Harmany, R. Marcia, and R. Willett, “This is SPIRAL-TAP: Sparse poisson intensity reconstruction algorithms - theory and practice,” *IEEE Trans. Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[6] S. Sra, D. Kim, and B. Scholkopf, “Non-monotonic poisson likelihood maximization,” Max Planck Institute for Biological Cybernetics, Tech. Rep. 170, 2008.

[7] R. Willett, “The dark side of image reconstruction: Emerging methods for photon-limited imaging,” https://sinews.siam.org/DetailsPage/tabid/607/ArticleID/220/The-Dark-Side-of-Image-Reconstruction.aspx, accessed: 20 Jan 2016.

[8] S. Gleichman and Y. Eldar, “Blind compressed sensing.” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6958–6975, 2011.

[9] A. Rajwade, D. Kittle, T.-H. Tsai, D. Brady, and L. Carin, “Coded hyperspectral imaging and blind compressive sensing.” *SIAM J. Imaging Sciences*, vol. 6, no. 2, pp. 782–812, 2013.

[10] S. Lingala and J. Mathews, “Blind compressive sensing dynamic MRI,” *IEEE Trans. Med. Imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.

[11] M. Collins, S. Dasgupta, and R. Schapire, “A generalization of principal component analysis to the exponential family,” in *Advances in Neural Information Processing Systems*, 2001, pp. 617–624.

[12] P. Hoyer, “Non-negative sparse coding,” in *Neural Networks for Signal Processing*, 2002, pp. 557–565.

[13] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, “Poisson noise reduction with non-local PCA,” *J. Math. Imaging Vis.*, vol. 48, no. 2, pp. 279–294, 2014.

[14] R. Giryes and M. Elad, “Sparsity-based poisson denoising with dictionary learning,” *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5057–5069, 2014.

[15] A. Soni and J. Haupt, “Estimation error guarantees for poisson denoising with sparse and structured dictionary models,” in *IEEE International Symposium on Information Theory*, 2014, pp. 2002–2006.

[16] B. Zhang, M. Fadili, and J. Starck, “Wavelets, ridgelets, and curvelets for poisson noise removal,” *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1093–1108, 2008.

[17] M. Mäkitalo and A. Foi, “Optimal inversion of the Anscombe transformation in low-count poisson image denoising,” *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, 2011.

[18] D. D. Lee and S. H. Seung, “Algorithms for non-negative matrix factorization,” in *Neural Information Processing Systems*, 2000, pp. 556–562.

[19] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, “Single pixel imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.

[20] Y. Oike and A. E. Gamal, “CMOS image sensor with per-column sigma delta ADC and programmable compressed sensing,” *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 318–328, 2013.